**1. What does one mean by the term "machine learning"?**

Learning and thinking process on Data is called Machine learning.

& Machine learning is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on patterns and inference instead. Building a model by learning the patterns of historical data with some relationship between data to make a data-driven prediction.

**2.Can you think of 4 distinct types of issues where it shines?**

Machine learning algorithms have had good results on problems such has spam detection in email, cancer diagnosis, fraudulent credit card transactions, and automatically driving vehicles.

**3.What is a labelled training set, and how does it work?**

A labelled training set is a collection of data where one of the features of the data indicates the class the training example belongs to. A labelled training set is used in supervised learning algorithms.

**4.What are the two most important tasks that are supervised?**

The two most common supervised learning tasks are regression and classification. In a regression problem we our prediction is a scalar value. When we're trying to solve a classification problem, our output is either 1 or 0.

**5.Can you think of four examples of unsupervised tasks?**

Common unsupervised tasks include clustering, visualization, dimensionality reduction and association rule learning.

**6.State the machine learning model that would be best to make a robot walk through various unfamiliar terrains?**

I would use a reinforcement learning approach. Reinforcement learning is a system where an "agent" observes the environment, selects and performs actions, then receives a reward or punishment based on the result of the action. Over time the agent learns by itself what is the most productive strategy.

**7.Which algorithm will you use to divide your customers into different groups?**

I would use some sort of clustering algorithm that can find the decision boundaries in the groups automatically. This is an unsupervised approach. However, if I already knew the categories of my customers, then I would choose a supervised approach and go with a classification algorithm.

**8.Will you consider the problem of spam detection to be a supervised or unsupervised learning problem?**

I would frame it as a supervised learning problem because humans have a general idea about what spam is and what it isn't. We can use this notion to create a labelled dataset for an algorithm to learn from.

**9.What is the concept of an online learning system?**

An online learning system learns from new data on-the-fly. As a result, the system is trained incrementally either by using one example at a time or using a mini-batch approach. This keeps each learning step cheap and memory efficient.

**10.What is out-of-core learning, and how does it differ from core learning?**

Out-of-core learning is used when a dataset is too large to fit into a computer's memory. The algorithm loads part of the data, runs a training step, then repeats the process until it has run on all the data.

**11.What kind of learning algorithm makes predictions using a similarity measure?**

Learning algorithm that relies on a similarity measure to make predictions is instance-based algorithm.

**12.What's the difference between a model parameter and a hyperparameter in a learning algorithm?**

Model parameter determines how a model will predict given a new instance; model usually has more than one parameter (i.e. slope of a linear model). Hyperparameter is a parameter for the learning algorithm, not of a model.

**13.What are the criteria that model-based learning algorithms look for? What is the most popular method they use to achieve success? What method do they use to make predictions?**

Model based learning algorithm search for the optimal value of parameters in a model that will give the best results for the new instances. We often use a cost function or similar to determine what the parameter value has to be in order to minimize the function. The model makes prediction by using the value of the new instance and the parameters in its function.

**14.Can you name four of the most important Machine Learning challenges?**

Four main challenges in Machine Learning include overfitting the data (using a model too complicated), underfitting the data (using a simple model), lacking in data and nonrepresentative data.

**15.What happens if the model performs well on the training data but fails to generalize the results to new situations? Can you think of three different options?**

If the model performs poorly to new instances, then it has overfit on the training data. To solve this, we can do any of the following three: get more data, implement a simpler model, or eliminate outliers or noise from the existing data set.

**16.What exactly is a test set, and why would you need one?**

Test set is a set that you test your model (fit using training data) to see how it performs. Test set is necessary so that you can determine how good (or bad) your model performs.

**17.What is a validation set's purpose?**

Validation set is a set used to compare between different training models.

**18.What precisely is the train-dev kit, when will you need it, how do you put it to use?**

If you tune hyperparameters using the test sets, then it may not perform well on the out-of-sample data because the model is tuned just for that specific set.

**19.What could go wrong if you use the test set to tune hyperparameters?**

Cross-validation is a tool to compare models without needing a separate validation set. It is preferred over validation set because we can save from breaking of part of the training set to create a validation set, as having more data is valuable regardless.